Supplementary Figures for: Markov State Model Reveals Folding and Functional Dynamics in Ultra-Long MD Trajectories

Thomas J. Lane,[1] Gregory R. Bowman,[1] Kyle Beauchamp,[2] Vincent A. Voelz,[1] and Vijay S. Pande[1]

Department of Chemistry[1] and Biophysics Program[2], Stanford University, Stanford, CA 94305

References

The following are the complete references for those with authors cited as "*et. al.*" in the manuscript.

8. Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, R. C.; Ierardi, D. J.; Kolossvary, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. *ACM SIGARCH Computer Architecture News* **2007**, *35,* 1.

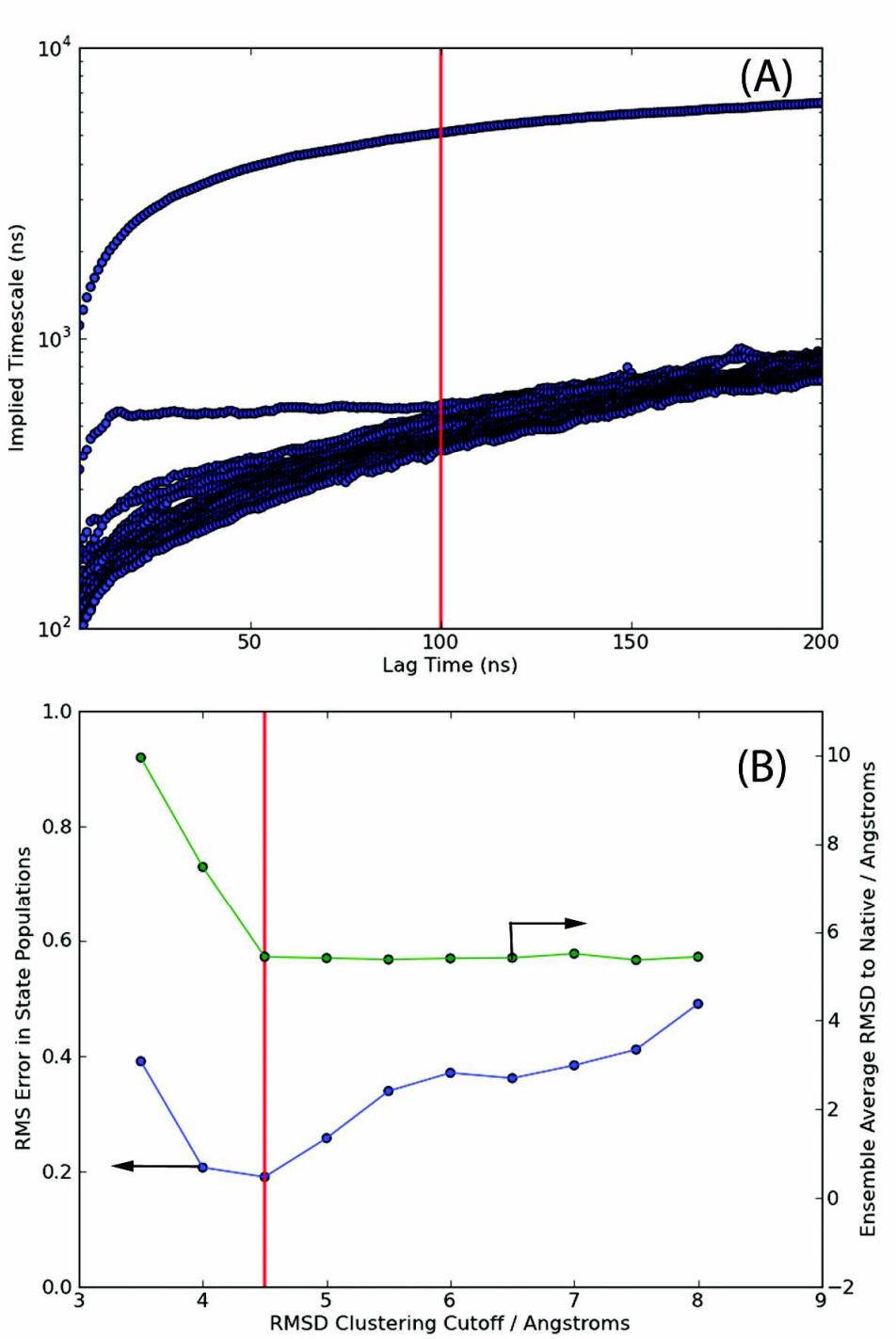

Figure S1. Implied timescales and equilibrium properties used to select model. (A) Implied timescales as a function of model lag time (chosen model: red line). Invariance in the implied timescales as the lag time is increased is a test of the Markovian properties of the underlying dynamics. (B) Equilibrium properties as a function of clustering cutoff. Show is the RMS error in state populations between the model and raw data (blue) and the ensemble average RMSD to the native state (green), chosen model is the red line.

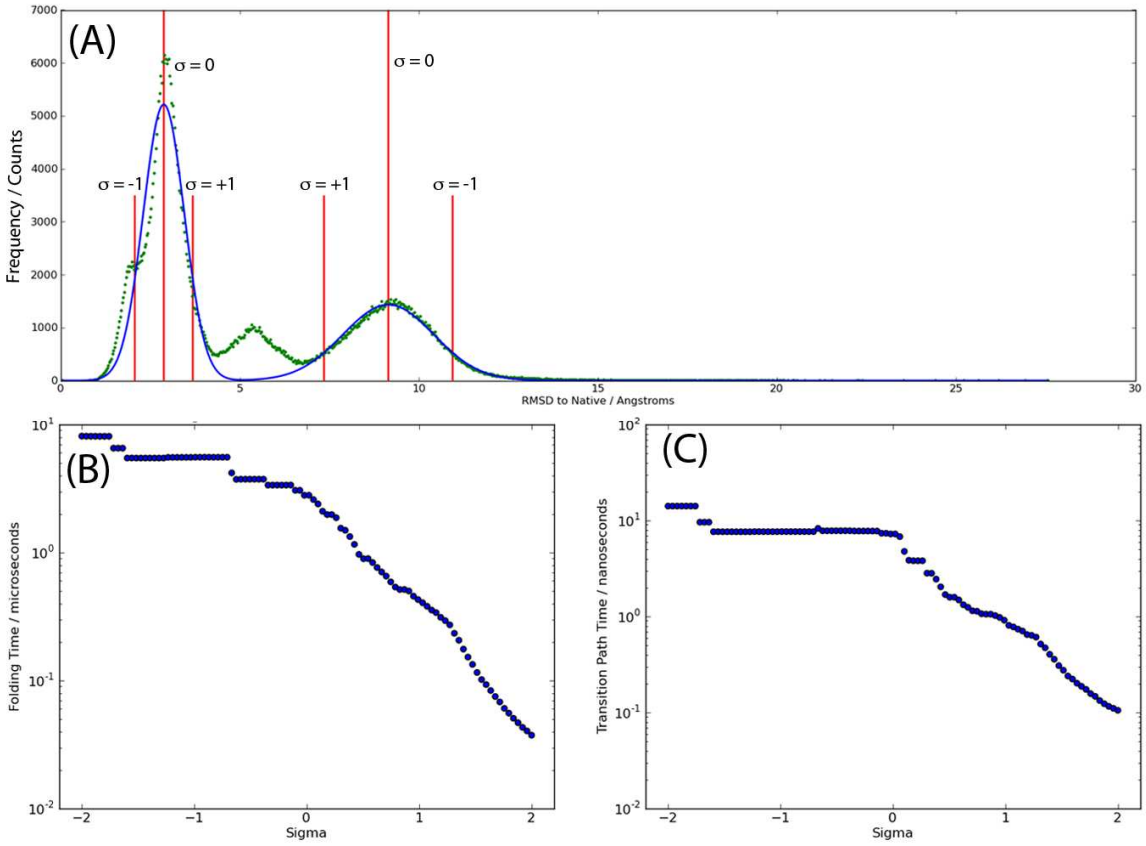

Figure S2. Variance of the (B) folding time and (C) transition path time with respect to state definition, as defined by (A) a two-state model created by Gaussian fitting. The red lines in (A) show the means of U and N, which would also be the cutoff at a *sigma* of zero in (B) and (C). The value of *sigma* then represents a fraction of one standard deviation of the Gaussian fit. The resulting fits span two decades of timescales as state definition is changed. This demonstrates that state definitions are essential to calculations of the folding time.

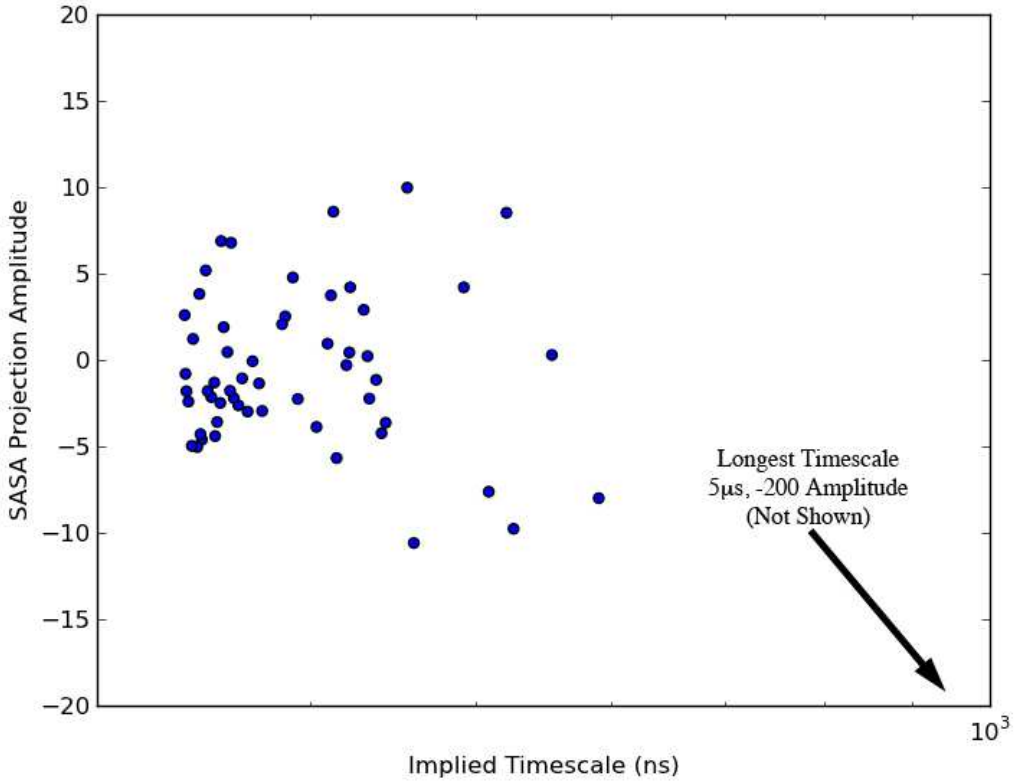

Figure S3. Projection of the eigenvectors onto the Trp8 SASA observable. The implied timescale gives a rate, while the amplitude gives a measure of the magnitude of change in signal at that rate. Seen are many processes that contribute to the bi-exponential relaxation in Figure 2. The longest dynamical process, with amplitude -200, is not shown so as to allow clear scaling of the plot.

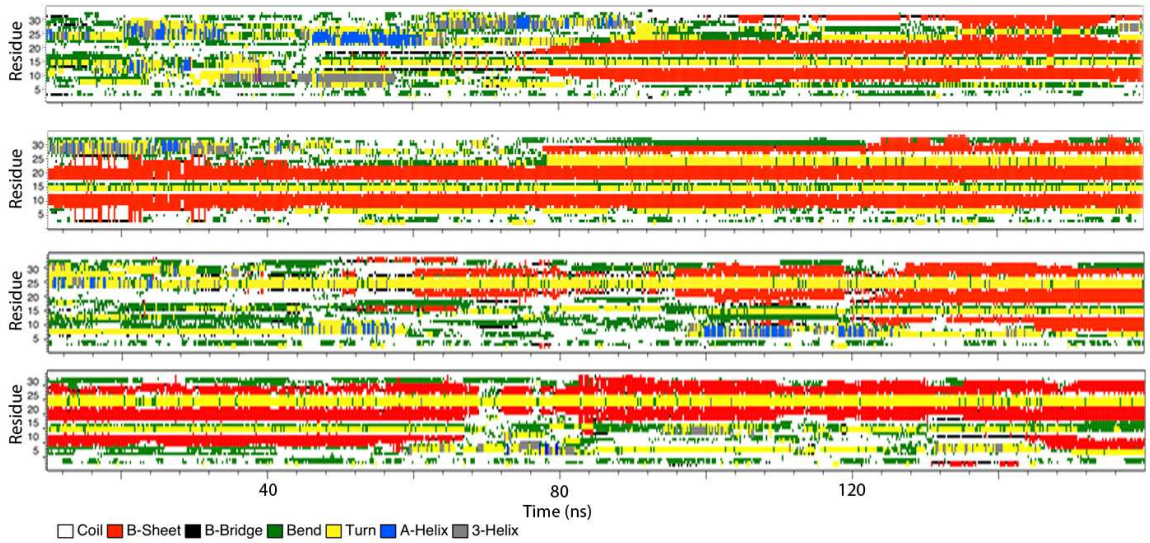

Figure S4. DSSP of the raw data over four 150 ns sections of the raw MD trajectories. Clearly shown in both cases are folding events involving the formation of either hairpin 1 (top two traces) or hairpin 2 first (bottom two traces). The red bands indicate beta-sheet structure, the bottom and middle bands denoting hairpin 1, while the top and middle bands depict hairpin 2.